International Journal of Civil Engineering and Technology (IJCIET)
Volume 8, Issue 3, March 2017, pp. 441–456 Article ID: IJCIET_08_03_045
Available online at http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=3
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

THE STATISTICAL ANALYSIS OF ACCIDENT-FREQUENCY DATA: A CASE STUDY OF NATIONAL HIGHWAY NO 6

Dr. V. S. Landge
Department of Civil Engineering
Faculty of Transportation Engineering,
Visvesvaraya National Institute of Technology Nagpur,
Nagpur-440010, Maharashtra, India

Koushik Kalita
Department of Civil Engineering
M. Tech. Student, Transportation Engineering
Visvesvaraya National Institute of Technology Nagpur,
Nagpur-440010, Maharashtra, India

Bahuguna Dalai
Department of Civil Engineering
PhD. Student, Transportation Engineering
Visvesvaraya National Institute of Technology Nagpur,
Nagpur-440010, Maharashtra, India

ABSTRACT

Road accidents in India have risen to it toll in the last few decades. Consequently, researchers in various part of the world have settled on the discussion of traffic safety management factor. An accident model has to be developed to establish the factors that cause the accidents. Accident model aids in determining the real causative agents of the accidents being witnessed across the country. The research article develops two accident models which are to be applied in various roads based on the location of the roads in the country such as in rural and urban roads. The main models which are mainly used in both rural and urban areas include linear and binomial distribution models. The data generated by several researchers indicates that the accidents in urban areas are attributed to age, speed, and gender. On the other hand, the over-speeding in rural roads has been noted the main cause of accidents (Ivan et al. 2000; Sumer 2003) [2, 12].

Vehicles, road conditions and human beings have an influence on the management of road safety. Investigation to establish the relationship between the accidents cases and the design of the roads has to be conducted. The main aim of the article is to establish the relationship between the accidents rates and the parameters employed in

the design. The article also develops models which are used in predicting accidents in both rural and urban roads (Mohan et al. 1985) [7].

Key words: Accident Rates, Regressions Models, Negative-Binomial Distribution and Poisson distribution

Cite this Article: Dr. V. S. Landge, Koushik Kalita and Bahuguna Dalai, the Statistical Analysis of Accident-Frequency Data: A Case Study of National Highway No 6. *International Journal of Civil Engineering and Technology*, 8(3), 2017, pp. 441–456.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=3

1. INTRODUCTION

Road accidents are the main cause of death, injury and disability in the world in the 20th century. Road accidents cause physical, financial and mental trauma to the affected families. India has recorded the highest road accidents related deaths with over 140,000 death persons annually. It is estimated that a total of 14 people lose their lives every hour. The research conducted by several researchers indicates that most accident cases are caused by human factors. The drivers, pedestrians, vehicles and poor designed roads are to be blamed for most accident fatalities. Human factors are the main cause of accidents on various roads. The driver, however, should not be the only person who carries the blame for causing accidents. Blaming the driver alone will prevent the engineers from seeing other factors which are responsible for causing accidents (Lord et al, 2004) [5]. The ability to identify all the other causative agents of accidents helps in developing ways of preventing accidents. The accidents rates can be largely minimized if the road factors and design are managed properly. Traffic accident models help the engineers to predict the accidents which are attributed to human factors and poor road design (Mohan et al. 1985) [7].

Modernization, Industrial revolution, and urbanization have resulted in increase in road traffic in most roads. The vehicles caught up in the traffic causes a lot of deadly accidents during rush hours of the day. Road traffic is a common phenomenon in rural roads in India. The traffic is caused by fast moving vehicles such as cars and buses and the slow moving vehicles such as carts and pedestrians. India recorded approximately 134,000 deaths caused by accidents in the year 2010 alone. 12,204 of the deaths were the pedestrians, and this formed 9.1% of the total accident deaths.4,000,000 pedestrians are killed annually in the world due to road accidents. The pedestrians have become susceptible to road accidents as the numbers of automobiles being used in roads increase (Parida et al.2005) [8]. The pedestrian deaths are common in countries with poor law enforcement such as India. Pedestrians are rarely considered when designing the road system despite the fact that they are the legitimate road users. Everyone becomes a pedestrian sometimes in a lifetime. Therefore, the pedestrian should not be overlooked in the transport matters. Most of the road planners involved in the design of roads gives little concern to pedestrians. Roads are being widened every day to accommodate large trucks without leaving space for the pedestrians. Pedestrian accidents can be minimized if the safety managers and road planners take into consideration the lives of the pedestrians. The developing an efficient traffic prediction model is paramount in the prevention of accidents and designing of safe and comfortable roads to the users(Mohan et al. 1985) [7].

2. LITERATURE REVIEW

A lot of investigations have been conducted to establish the causes of accidents, but the article analyses the development of mathematical models which can predict road accidents. A generalized linear regression model was the first model to be developed to predict road accidents. The model used in the prediction of accidents on two-lane rural roads. Notably, several types of research which have been conducted documented reports which analyzed geometrical design standard and how it affect road accidents, analysis of traffic and accident prevention.

Firstly, Poisson-Gamma and Poisson-Weibull model were used to analyze road accidents on National Highway (NH) number 58 in India. The research indicated that commercial, residential and schools were responsible for causing accidents. Other causes of the accidents identified by the researcher include road side investment developers, vehicles traffic, and median opening. Secondly, Regression analysis was employed in studying accidents occurrence in NH-77.The outcome of the research was to be used in developing mathematical accident model. The findings of the research revealed that the accidents are attributed to AADT and the condition of the road. Thirdly, Multiple Linear Regression was used to predict the number of accidents in Bhopal city. The results of the research gave cross-section dimensions, traffic volume, and traffic signs as the main causes of accidents. Other causes of accidents identified by the research include the width of the road shoulder, conditions of lightings and traffic lights. Another research conducted using multiple linear regressions indicated that the accidents are attributed to the number of junctions, horizontal and vertical curvatures. The causes of accidents included surface irregularity and road width in Kenya and Jamaica respectively. Moreover, multiple linear regressions types of research which were conducted differently in various roads identified access point and approach speed as the major causes of accidents. Vehicle traffic and the gap were also found to be the causative agent(Ivan et al. 2000) [2].

. Another research pointed speed, motorists, and pedestrians as the major causes of accidents. Additionally, the negative binomial multivariable analysis was used to develop a model in Spain. The analysis was conducted on two-lane rural roads. The causative agents established in the research include access density, approach speed, and sight distance. Access density refers to a variable which is responsible for most of the head-on collision in vehicles (Mohan et al. 1985) [7].

The urban road which is being analyzed in the article is the four-lane National highway (NH-6) in Maharashtra. The model to be used in the urban road was developed by the use of multiple linear regression analysis. The accidents recorded in the highway are attributed to the design parameters of the road such as road roughness and as carriageway width. The model developed to measure the accidents and accident related factors in the rural roads are the binomial distribution model. The models developed are essential in the design of safe and comfortable roads in the country. The models will also aid in determining the causative agents of accidents in the main roads so as to devise the methods of averting them. Well-developed models guarantee a decline in road accidents and hence a reduction in the loss of lives ((Sümer 2003) [12].

3. URBAN ROAD

3.1. Methodology

The dependent variable used in the study is the accident rates while the width of the road, Horizontal Curvature, the roughness of the road and Vertical Curvature represent the independent variables. The other elements which have been considered as independent variables include sight distance, the number of junctions and Service Road.

The research was conducted on an eighty-two kilometers road section. The data was collected in the between 2015 and 2017. The data collected was divided into two parts. The first part was used in the developing traffic Prediction model. The second part of the data was used in validating the model (Karlaftis et al. 2002) [4].

3.1.1. Calculation of Accident Rates in the road section

$$A_R = \frac{C*100000000}{V*365*N*L}$$

A_R = road accidents occurring in every one hundred million vehicles traveling in the road segment per year

V =quantity of traffic in the road segment

L = length of the road section

N =the periods the data has been collected in years.

C = road accidents recorded in the road segment.

Homogenous sections

The road segment being studied was divided into homogeneous sections depending on the pattern of the accidents observed. The creation of the homogenous section was done using the technique of cumulative difference area. The idea behind the Cumulative difference is derived from the statistic Z_X which stands for the difference between cumulative area and mean areas under the drawn curve. The formula below gives how the value of $Z_X$is calculated (Author unknown no year).

$$Z_{x=} \sum_{i-1}^{n} a_i\text{-}\left(\frac{\sum_{i-1}^{n} a_i}{L_p}\right)\sum_{i-1}^{n_i} x_i \ ; a_{i=\left(\frac{r_{i-1}+r_i}{2}\right)} x_i$$

L_P=length of the road segment

n_t= the number accidents that occurred in the road segment

r_i= the forms of accidents on the road

In geometry,Z_X is given by (cumulative are $+r_i$)-(mean cumulative area) in the same distance of the road segment being studied (Karlaftis 2002) [4].

3.1.2. Multiple Linear Regressions Analysis

For the purpose of predicting the dependent variable which is the accident rate, the dependent variables are identified first so that the random sample values of n-size of the dependent variables can be found (Joshua et al. 1990).

Assuming K is the no of independent variable:X_i, i = 1, 2... K, while Y is assumed to be the random variable which is dependent, then the equation below gives the multiple linear regression formulae.

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_K + \varepsilon_i \quad (i)$$

Y_i and X_1, X_2, X_K represent the dependent random and independent variables respectively while, β_O, β_1,., B_K represents parameters of the model whileε_i represents the random error which exhibits a normal distribution. $\varepsilon_i$has both zero mean and its variance is always constant.

Multiple linear regression models are made up of two parts which are a determined (Y'_i) and stochastic (ε_i).

Yi' = β_0 +$\beta_1 X_1$ +$\beta_2 X_2$ + ...+$\beta_K X_K$ (ii)
From equation (i)

ε_i= $(Y_i - Y'_i)$ (iii)

The average of the dependent variable (Yi) is done for a given independent variables forms the determined part.

Y'_i = E (Y'_i)

=Y_i =β_0 +$\beta_1 X_1$ +$\beta_2 X_2$+…+ $\beta_K \beta_K$ (iv)

The elements shown by Y_i are the average of E (Y_i). The model given by equation (i) was developed by estimating the samples regression model.

ŷi=b_0+$b_1 x_1$+$b_2 x_2$+….+$b_k x_k$ (v).

ŷi represents dependent variable value which has been adjusted.

Y_i, X_1, X_2, X_K represents the values of independent variables andβ_O, β_1,., B_K represents the values which have been estimated β_0.

The traffic accidents forecasting models which should be chosen should give the relationship of the pattern observed effectively. The objective will be achieved by making sure that the sums of squares equations are minimized. The regression plane, for instance, where the value of k is 2 should be minimized.

$\sum e_i^2$=$\sum(y_i$-$\hat{y}_i)^2$=minimum (vi)

e_i Is the random error in the sample.The model not only focuses on representing the mathematical equations but, the model is also based on some assumptions which necessitate the use of the estimated values of β_O, β_1,., B_K.The conventions used in the development of the model are related to the random error. The random error depicts normal distribution behavior which is equal to zero. All the values which provide support and contained in the random error have equal variances. When the value of k is 2, the regression plane equation being derived is the simplest form as shown below (Joshua et al. 1990) [3].

ŷi =b_0 + $b_1 x_1$ + $b_2 x_2$ (vii)

The standard error samples in the study are used to determine the version of the estimated regression model adapted from the use of the empirical data. The standard error used encompasses the estimated value of standard deviation encountered in the tabulation of random error σ_ε. The standard error is represented as Sε which is the square root of repetition .the equation below gives how is calculated.

R^2=$\sqrt{\frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\hat{y}_i)^2}} = \frac{SSR}{SSy}$ (viii)

The independent variable is given by the following expression

R_{adj}^2=1-$\frac{n-1}{n-k-1}(1-R^2)$ (ix)

n represents the size of the sample while k represents the total number of independent variables (Joshua et al. 1990) [3].

3.2. Testing of the Model Usage

The significance of the given estimates is tested before the estimated regression equation is put into practice. Consequently, a hypothesis which depicts zero as well alternative values is employed (Miaou 1996) [6].

H_0:$\beta_0 = \beta_1 = \beta_2 = \ldots + \beta_K = 0$

H_A: at least one$\beta_i \neq 0$

Zero hypotheses have been laid to ensure that existence of a linear relationship between variations of phenomena which are observed is not exhibited. The connection has no relationship when X_1, X_2, …, X_K does not have influence on the value of Y. The change independent as a result of the change in independent variables is given by

SSy = SSR + SSE (x)

SSy represents the total variability, SSR stands for explained variability, and SSE represents unexplained variability. F-test is applied to test the chances of regression model working while employing the use of analysis of variance (Miaou 1996) [6].

Table 1 shows how the analysis is done.

SR	DoF	SoS	MS	Ratio of F
Regression	k	SSR	MSR=$\frac{SSR}{k}$	
Errors in the sample	n-k-l	SSE	MSE=$\frac{SSE}{n-k-1}$	F=$\frac{MSR}{MSE}$
Sum	n-1	SSy		

SR=sources of variation

DoF=Degrees of freedom

SoS=Sum of squares

MS=mean square

The following rules are applied while making the decisions.

If F ≥ Fα; k; n-k-1, indicates that the hypothesis is null, and it is discarded but if F < Fα; k; n-k-1, the hypothesis is never rejected .the hypothesis indicated that the values of independent variable do not affect the values of dependent variables.

Testing of Regression Coefficients

The implication of the estimates of the parameters(β_i, i=1, 2, … k) is determined if the estimated regression model is to be employed in predicting the dependent variable Y.in the instance where k is taken to be 2 the estimates of β1 and β2 is tested(Miaou 1996) [6].

H_0:$\beta_0 = 0$ (2) H_0:= $\beta_2 = 0$

H_A: $\beta_1 \neq 0$ β_2: $\beta_2 \neq 0$

The following statistical analysis is done

$t_1 = \frac{b_1}{s_{b1}}$

$t_2 = \frac{b_2}{s_{b2}}$

The statistic has t-distribution, and the freedom degree is given by n – k – 1.

$|ti| < t\alpha/2$, i=1, 2, makes the hypothesis to be accepted. In this case, the there is no influence of an independent variable on dependent variables. The following test is always done when using multiple regression models (Miaou 1996) [6].

H_0: β_1= 0

H_A: $\beta_1 \neq 0$

i = 1, 2…… k

Statistical analysis is given by

$$t_i = \frac{b_i}{s_{bi}}$$

The test has a t-distribution while n – k – 1 with degrees of freedom of the equation. The null hypothesis will only be accepted if $|ti| < t\alpha/2$(Joshua et al. 1990) [3].

The process involved in the development of a multiple regression model and testing it needs a lot of mathematical calculation to be done. The hypothesis has to be tested to ascertain that the value calculated for every regression coefficient arose due to computation that was done. The results of the test will be checked by computing the standard error each regression coefficient, and then it tested 5% level to establish its significance. The variables which do not exhibit any significance are discarded. The Microsoft Excel in computers is used to remove manually. The variables with no significance are eliminated out of the analysis when using multiple regression analysis. The Microsoft Excel also tests some variables for significance and replaces some variables when they are worth replacing. The technique of using the Microsoft Excel to test variables and to eliminate the non-worthy variables is known as stepwise regression analysis (Peng et al 2014) [10].

3.3. Data Collection

The NH-6 under investigation runs through Bhandara , Nagpur and Amravati districts of Maharashtra. The geographical location of Nagpur is 21.1458°N and 79.0882°E. The highway is elevated 310 m above sea level. The road passes an area which is marked with open and rolling terrain all through its length. The road also encounters some hilly terrains in some parts of its length. The road connects Kolkata and Surat.

3.3.1. Homogenous Sections

Table 2 Length of homogenous sections under study

Sec. No.	From Km	To Km	Total length Km
1	638.0	647.0	9.0
2	647.0	657.0	10.0
3	659.0	664.0	5.0
4	666.0	670.0	4.0
5	670.0	672.0	2.0
6	672.0	681.0	9.0
7	681.0	692.0	11.0
8	694.0	697.0	3.0
9	697.0	704.0	7.0

Table 3 Frequency of accidents occurrences

Sec. No.	Stretches Km	Total Length Km	Accident Rates (Yr-2013)	Accident Rates (Yr-2014)	Accident Rates (Yr-2015)	Accident Rates (Yr-2016)	Accident Rates (Yr. 2017)
1	638.0-647.0	9.0	91.776	80.530	91.776	86.153	88.027
2	647.0-657.0	10.0	24.137	55.559	32.421	29.279	46.132
3	659.0-664.0	5.0	19.031	12.354	25.708	12.354	19.031
4	666.0-670.0	4.0	39.062	92.479	39.062	39.062	56.868
5	670.0-672.0	2.0	12.354	19.031	12.354	12.354	14.580
6	672.0-681.0	9.0	59.700	69.412	69.412	57.272	66.175
7	681.0-692.0	11.0	21.257	34.611	47.965	14.580	34.611
8	694.0-697.0	3.0	81.125	47.965	56.868	61.319	61.319
9	697.0-704.0	7.0	86.755	48.601	56.232	33.339	63.863

3.3.2. Road Design

Table 4 Various geometric parameters of the road sections

SN	AWR(m)	VC(m/km)	HC(degrees/km)	RI(mm/km)	NJ/km	SD(m)	SR (%)
1	17.81	15.66	32.29	1881	2.83	273	93
2	17.92	22.81	43.39	2054	1.41	268	98
3	18.02	6.76	10.44	1926	1.24	289	99
4	17.9	22.59	56.81	2162	1.99	269	99
5	14.94	14.68	33.88	2220	2.12	261	76
6	15.12	12.87	20.94	1909	2.81	274	94
7	15.21	8.93	23.84	1947	1.24	290	89
8	16.20	17.16	56.78	2126	1.49	248	69
9	14.33	17.16	25.21	1956	3.42	267	88

AWR=average width of the road
VC=vertical curvature
HC=horizontal curvature
RI=rougness index
NJ=number of junctions
SI= sight distance
SR=Service road

3.3.3. Site Selection

The highway under study stretches between 638 km to 704 km. The total distance to be studied 82km.

3.3.4. Traffic Volume (AADT)

Both the primary and the secondary data was collected by the guidance given in IRC SP 19-2001.The data obtained was subjected to detailed analysis to give ADT information and AADT. The AADT was taken in the year 2017.The information was taken throughout the period of study.

3.3.5. Accident rates in the road section

The accident data was recorded on the road between 2013 December and 2016 December. The data was obtained between Amravati and Nagpur national highway No.6.The accident rates in every one hundred million vehicle-Km of travel in a year was determined through calculation.

3.3.6. Horizontal Curvature

The expression below is used to calculate the horizontal curvature. Average curvature degree between point A and B $=\frac{Q1+Q2+Q3\ldots\ldots Qn}{distance\ AB}$.

The units of measurement are degrees/kilometer.

3.3.7. Vertical Curvature (VC)

Rise and fall measured per kilometer are known as vertical curvature.

3.3.8. Sight Distance

Sight distance refers to a minimum actual distance on the road in which the driver has an ample time to stop without causing a collision in case the vehicle was traveling at high speed. The approved sight distance for a vehicle traveling at 100kmph on a four-lane highway is 360m.the sight distance that was calculated for the road in study basing on the regulations of IRC 73-1980 & IRC SP 23-1993 are shown in the table below.

Table 5 Regulation to calculate sight distance as per IRC 73-1980 & IRC SP 23-1993

SN	SD	EHD	OH
1	SSD	1.15	0.15
2	ISD	1.15	1.15
3	OSD	1.15	1.15B

SD=sight distance

EHD=eye height of the driver

OH=object height

SSD= sight distance of stopping

ISD= sight distance which is intermediate

OSD=height distance of vehicle overtaking

3.3.9. Number of Junctions (JN)

The junctions in the road in the study were counted on both sides of the road. The junctions in the crossroads were considered as two junctions while the junctions on a five-way road were

considered as to be three junctions. Both the slip and slip out was counted as a junction. The junctions were expressed as in junctions/km.

3.3.10. Carriageway width (CW)

The width of the pavement in the road segment was measured in meters and given in average width/km of the road segment.

3.3.11. Road Roughness (RI)

Roughness Index gives the features of the longitudinal profile of the road segment. The standardized unit of measurements is used to establish the roughness of the road. The unit of measurement of roughness index is mm/km.

4. ANALYSIS AND DISCUSSION OF THE RESULTS

The connection between the dependent and independent variables is given by the expression below.

Y = b_0 + b_1X_1 + b_2X_2+b_3X_3+ b_4X_4+b_5X_5 + b_6 X_6+... ... + b_nX_n and Y,X_1, X_1.....$X_n$are dependent and independent variables respectively ,b_0 represent regression constant while b_1, b_2.....b_n areRegression coefficient .The traffic road accident model was developed using the data in the table below.

Table 6 Data used to develop the traffic road accident model

SN	AG	R_w	V_C	H_C	$\frac{R_1}{R_W^2}$	J_N	SD	SG
	(Y)	X_1	(X2)	(X3)	(X4)	(X5)	(X6)	(X7)
1	89.02	17.81	15.66	32.29	5.91	2.83	273	93
2	47.13	17.92	22.82	43.38	6.38	1.40	268	98
3	20.03	18.02	6.76	10.44	5.92	1.24	289	99
4	57.86	17.90	22.59	56.81	6.67	1.99	269	99
5	15.58	14.94	14.68	33.88	9.93	2.12	261	76
6	67.17	15.12	12.88	20.94	8.33	2.81	274	94
7	35.61	15.22	8.93	23.84	8.39	1.24	290	89
8	62.31	16.20	29.26	56.78	8.08	1.49	248	69
9	64.86	14.33	17.16	25.21	9.50	3.42	268	88

5. RESULT & VALIDATIONS

5.1. Results

An equation which gives a relationship between accident rates in every 100 million travel annually and design parameters was developed using multiple regression analysis. The outcome of the analysis indicates how the captured factors cause accidents. The equation developed shows how the elements are related to accidents as shown below.

$$A_R\text{=}867.77\text{-}35.76R_W\text{+}1.23H_C\text{-}40.45\frac{R_1}{R_W^2}\text{+}19.13J_N \qquad (1)$$

The model developed indicates that it provided a good fit on the side of the independent variable given that R^2 is 0.9873 and the value of adjusted R^2 is 0.9746 which is significantly large .The percentage of R^2, which represents the total number of variations in recorded accidents is 98.73% and its value of probability is 0.00048 .the probability for the

independent variables was found to be less than (0.05).the results show that the relationship is only significant to 5.0% level.

Table 7 Result of ANOVA test

ANOVA						
		DoF	**SoS**	**MS**	**F**	**SIGN.F**
Model	regression	3.99	4482.93	1120.73	77.76	0.0005
	residual	3.99	57.65	14.41	77.76	
	total	7.99	4540.58	1135.14	155.52	0.0005

6. VALIDATION OF MODEL

The following section provides the procedure that was used in validating the prediction model that was developed. The validation will be conducted by analyzing the differences noted in the predicted values. The two elements below were used in the validation of the developed model (Miaou 1996) [6].

6.1. Validation of the Models by the use of data from a Different Study Year in the Same Road Segment.

The validation is used to establish whether the model can predict accidents at all time. The data to be used in validation was collected in 2011 from the same road segment which stretches between 638km to 704km.Theaccident rates was calculated using Equation -1.The forecasted accident rates for the year are shown in the table below.

Table 7 The forecasted accident rates for the year 2011

NO.	Km	RW	HC	$\frac{\boldsymbol{R1}}{\boldsymbol{R_W^2}}$	JN	ACCIDENT RATES	MAD	MAPE
1	638.5-655.0	17.81	32.20	5.91	2.83	84.86	12.80	44.5
2	655.5-660.5	18.02	1O.44	5.92	1.24	48.45		
3	660.5-665.5	14.94	33.88	9.93	2.12	13.40		
4	665.5-688.5	15.22	23.84	8.35	1.24	36.56		
5	688.5-704.0	15.74	14.28	8.33	4.13	63.59		

The absolute error (MAPE) calculated in the analysis in the observed and a forecasted value was found to be 45% while the (MAD) in the analysis was found to be low.

6.2. Validating the model by adding another study segment on the road.

The method is used to establish whether the developed model can predict accidents from a road segment whose data was not used in the development of the model.

In our study, the data used was extracted from the road segment which stretches between 690 km to 704 km. The accidents rates were forecasted using equation 1.the results of forecasting are shown below (Miaou et al. 1996) [6].

Table 8 The results of forecasting of accident rates

YEAR	DISTANCE	RW	HC	$\frac{R_1}{R_W^2}$	JN	PREDICTED ACCIDENT RATES	OBSERVED ACCIDENTS RATES	PERCENTAGE VARIATION
2015	690.5-704.0	15.74	14.28	8.33	4.13	63.59	47.90	-32.9
2016	690.5-704.0	15.74	14.28	8.33	4.13	63.59	51.51	-0.98
2017	690.5-704.0	15.74	14.28	8.33	4.13	63.59	53.42	-18.9

The variations in the observations made were found to be small.

7. A MODEL FOR RURAL ROADS

Road network in India is largely made up of rural roads. Most of the rural roads are poorly designed. The rural roads are mainly used by farmers to transport farm produce to the markets. The roads are potholed, and of low quality, therefore, a lot of accidents have been recorded which involve farm machinery equipment. The rural road under investigation is Berasia to Semrakalan Approach road in India (Persaud 1994).

7.1. Development of the Model

The number of accidents that are used in the research was recorded in a period of one year. The accidents are given a random variable Y .the prediction of the accidents in the road will employ the use of Poisson law.

P(Y=K) $=\frac{\lambda^k}{k!}e^{-\lambda}$

The parameters used are l>0 ,k =0,1,...,¥

The expected value of E(Y) is 1 while the dispersion D(Y) = l

Notably, it has been established that the dispersion is normally greater than the mean value when the data on accidents occurrence is being investigated. A phenomenon known as over-dispersion is exhibited. In the case where the application of Poisson distribution law appears to be dubious, another law which forecast road accidents has to be used. Let γ be considered as a random variable which is denoted as Δ.if the parameters are assumed to be distributed by Gamma distribution law, the parameters will assume the probability density function shown below(Persaud 1994) [11].

F(x)=$\frac{(\frac{\alpha}{\mu})^{\alpha}}{\alpha}x^{\alpha}e^{\frac{\alpha}{\mu}x}$=0

P(Y=K/∇=γ) $=\frac{\gamma^K}{K!}e^{-\gamma}$, k=0, 1,∞

The distribution of Y is then given by the following example.

P(Y=K) $=\int_0^{\infty}\frac{\gamma^k}{k!}e^{-\gamma}\frac{(\frac{\alpha}{\mu})^{\alpha}}{\alpha}\gamma^{\alpha-1}e^{\frac{\alpha}{\mu}\gamma}\,d\gamma$

The aim of the research is to forecast the number of accidents (random variable Y) basing on the characteristics of the road, the state of the drivers and the road traffics. The road segment to be studied is given by it features which include geometry, curves, road sign a speed limit. Every accident that occurs takes place within explanatory variables which are their function. The random variable Y will only be determined by establishing its parameters which are denoted by a and m .given that parameter can be influenced by the road conditions, the parameter m = E (L) need to be determined (Wang, Lee, et al 2003). The parameter m has been estimated while considering the road condition using the following model.

$$\mu = a_0 X_2^{a_1} X_2^{a_2} \ldots\ldots\ldots X_M^{a_M} e^{\sum_{e^{i=1}}^{N} b_1 Z_1} \quad (2)$$

X_1,...X_M, Z_1, ...Z_N= explanatory variables whilea_0 ,...a_M,b_1 ,...b$_N$are the coefficients which have not been determined.

7.2. Model Design

The choice explanatory variables should be based on the data extracted from the rural road. The road under investigation is divided into sections .Each section has its characteristics which are used to form explanatory variables. The total numbers of accidents which occur in each road segment are recorded. The data is recorded over a long time to ensure that a large amount of data is recorded. The explanatory variables used in the research are shown below (Persaud 1994) [11].

Variable X_1which is known as seglenrepresent the road segment length

Variables Z_1 and Z_2 which are known as *speed* and *cross* respectively.

Z_1represent the speed limit, andZ_2 represent the total number of crossroads in the road segment.

The rural road under investigation was divided into 150 sections then 150 observation is made .Each road segment was assigned different explanatory variables then accidents on each section were recorded.

Table 9 The explanatory variables and no of accident data

Variables			Accidents	
Seglen	**Speed**	**cross**	**Observed**	**predicted**
0.79	70	3	3	2.0
1.12	60	6	9	8.5
0.56	80	2	2	1.1
1.1	50	6	7	10.3
1.2	50	4	5	6.6
1.3	60	3	4	3.3

Table 10 Regression Coefficient used and result of t-test

Regression Coefficient used		t-statistical	95 % Confidence level	
a0	0.843	5.44	0.54-1.15	Accepted
a1	1.400	13.92	1.20-1.60	Accepted
b1	0.08	3.57	0.04	Accepted
c1	0.32	15.88	0.28-0.36	Accepted

The observed numbers of accidents were tested by the used of Chi-square test. The results of the test are given below.

Table 11 Result of Chi-square test

	Frequency			Chi-square	
Accidents occurrence	**observed**	**Poisson**	**Negative binomial**	**Poisson**	**Negative binomial**
0-2	32	24.29	33.01	3.13	0.00
3	31	33.21	27.82	2.14	0.63
5	8	17.80	15.51	2.46	0.45
6	6	9.70	9.90	0.76	0.84
9-11	10	2.85	6.67	13.32	0.84

Variance = 4.799

Standard deviation = 2.1906

estimated value = 3.273

Estimated value of ∇of the Negative Binomial distribution = 3.273

Estimated of ∇of Negative Binomial distribution = 7.025

Chi-square test for Poisson distribution = 26.920

Chi-square test for Negative Binomial distribution = 6.789

Distribution Chi-sq. (26.920) = 1.000 > 0.95, D.f. = 6

Distribution Chi-sq(6.789) = 0.763 < 0.95, D.f. = 5

Hypothesis given by the Poisson distribution is discarded while the hypothesis provided by the Negative Binomial distribution is accepted only if the significance is 5%.

The equation 2 now is given as

$\mu = 0.843 X_1^{1.4} e^{0.08Z1+0.32Z2}$(3)

Equation 3given above can be used to forecast the number of accidents that are likely to occur on a given road in a given period. The number of accidents which can occur in a one-kilometer road with a fixed speed of eighty kilometers per hour is given as follows. (*seglen*= 1, *speed* = 6, *cross* = 3) Then the number of accidents is 4.2.the confidence level for this example is 95% confidence the interval is given as (2.6, 5.8).The number of accidents occurring on the road can be reduced to 3.6 and at a confidence level of 95% if the speed is limited to 60 km/hr.

8. CONCLUSION AND RECOMMENDATION

The research was done successfully, and accident traffic prediction model for urban and rural roads was developed. A multiple linear regression model is to be used in forecasting accidents in the rural highways while the binomial distribution model is to be used in forecasting road accidents in urban highways. The data applied included accident historical, geometric and facilities data. The model developed for urban roads indicates that the road parameters such as sight distance, vertical curvature, and service roads are not the main causative agents of road accidents on the highways. The road accidents in the highway were found to be caused by road design problems at 5.0% level. The design parameters that cause accidents at 5.0% level include road roughness and horizontal curvatures (Peden 2004) [9].

Road accidents are random and developing a model which predicts the exact number accidents in the given roads proves to be difficult to come up with. The models developed for rural and urban roads are of great importance in the design of safe roads. The models will also assist in identifying the dangerous locations on the roads. The locations will be improved to ensure safety once they are identified. The accidents are also caused by the human factors .the developed models, however, do not incorporate human factors in their functionality. The behavior of the driver and the modes of driving as a result of a change in traffic have not been considered during the development of the project. The inclusion of human factors in the initial stages of the development the model is paramount for an efficient traffic accidents prediction model (Peng et al. 2014) [10].

8.1. Recommendations

An extended research has to be conducted to validate the models on other highways and rural roads. The data generated by several researches shows that various models portray different behavior when they are used in different environments. Further research has to be done to validate the performance of the model in different types of highway facilities such as

expressway and intersections. The models need to be tested when various data sets are used. Moreover, an accident prediction model which puts into consideration the human factors needs to be developed (Peden 2004) [9].

Members of the public play important roles in maintaining safety in the rural and roads. The citizens need to be enlightened on the importance of the safety in the society. The move can be achieved by involving the members in road safety campaigns. The campaigns bring into attention the road safety guidelines in the road. The members of the public, for instance should be advised to use appropriate crossing areas such as the zebra crossing when crossing the road. The crossing of the road should be done only when the road is clear. The public get an opportunity also to make inquiries on pertinent issues that revolve around road safety. Most importantly, the public need to put into consideration road signs such traffic lights. The responsibility of guarding the public properties such as road signs should not be left to governmental agencies alone. The public should report the people who are involved in vandalizing road safety equipment. Areas with vandalized road safety equipment are prone to fatal accidents .the drivers should also be encouraged to embrace the spirit of improving road safety. The drivers need to observe the speed limit set in some particular places such schools and commercial centers. Most of the accidents which are normally witnessed in the rush hours of the day are attributed to drivers who fail to observe the speed limits in trading centers.

The road accidents can reduce significantly by using the developed accident prediction models. The models give necessary improvements which must be done to control safety on the roads. The crowded places should be provided with bypasses and fly-overs to reduce congestions and traffic when crossing the road a lot of junctions in the road have been linked to various. The road design and construction personnel should be instructed to reduce the number of junctions in the road. Too long flat roads force the drivers to speed up and omit the road safety procedures. The road engineers are advised to introduce the curves in the road and bumps to control overspending. Most of the accidents in the rural roads are caused by pedestrians and animals drawing farm produce. A pedestrian lane should be construction during the initial plan of the road to accommodate the pedestrians and the animals (Peden 2004) [9].

Additionally, the government should come up with legislations which are aimed at punishing the members of the public and vehicle owners who violate the road traffic rule and regulations. The perpetuators of the rules should be taken to court and judged accordingly. The old vehicles which compromise the safety of road users should be banned out of the roads. The drivers should pass through the required training before being allowed to operate on the roads. The drivers should also renew their license with the transport ministry to ensure that they meet the requirements for their work. The exits and entry sections of the road should be made clear to all the vehicles entering and exiting the section. The government should not allow traders to operate in proximity to the roads. The traffic accidents prediction models will help in curbing the increasing accident rates once implemented (Peden 2004) [9].

REFERENCE

[1] Ganveer, G. and Tiwari, R.Injury pattern among non-fatal road traffic accident cases: a cross-sectional study in Central India. *Indian journal of medical sciences,* 59(1),2005, pp. 9.

[2] Ivan, J. N., Wang, C. and Bernardo, N. R.Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis and Prevention*, 32(6), 2000, pp. 787-795.

[3] Joshua, S. C. and Garber, N. J.Estimating truck accident rate and involvements using linear and Poisson regression models. *Transportation planning and Technology*, 15(1), 1990, pp. 41-58.

[4] Karlaftis, M. G. and Golias, I.Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis & Prevention*, 34(3), 2002, pp. 357-365.

[5] Lord, D. and Persaud, B. N.Estimating the safety performance of urban road transportation networks. *Accident analysis and prevention*, 36(4), 2004, pp. 609-620.

[6] Miaou, S. P.Measuring the goodness-of-fit of accident prediction models. Report No. FHWA-RD-96-040, 1996.

[7] Mohan, D. andBawa, P. S.An analysis of road traffic fatalities in Delhi, India. *Accident Analysis and Prevention*, 17, 1985, pp. 33-45.

[8] ParidaM ,JainS.S. and Landge V.S.2005Modeling Traffic Accidents On Rural Two Lane Highway Under Mixed Traffic Conditions."

[9] Peden, M. World report on road traffic injury prevention, 2004.

[10] Peng, Y., Lord, D. and Zou, Y. Applying the Generalized Warning model for investigating sources of variance in motor vehicle crash analysis. *Accident Analysis and Prevention*, 73, 2014, pp. 20-22.

[11] Persaud, B. N.Accident prediction models for rural roads. *Canadian Journal ofCivil Engineering*, 21(4), 1994, pp. 547-554.

[12] I. Rohini and V. Arularasi, National Highway Design-A Case Study of Kadapa To Kurnool National Highway. International Journal of Civil Engineering and Technology, 7(4), 2016, pp.406–421.

[13] B.Naga Kiran, Dr. N. Kumara Swamy and Dr. C. Sashidhar, Prediction of Road Accident Modelling For Indian National Highways. International Journal of Civil Engineering and Technology, 8(1), 2017, pp. 789–802.

[14] Sümer, N. Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. *Accident Analysis and Prevention*, 35(6), 2003, pp. 949-964.

[15] Wang, K., Lee, A. H., Yau, K. K. and Carrivick, P. J.A bivariate zero-inflated Poisson regression model to analyze occupational injuries. *Accident Analysis and Prevention*, 35(4), 2003, pp. 625-629.